Le clonage positionnel : ou comment chercher une aiguille dans une meule de foin

*M. Gérard-Blanluet**

Le clonage positionnel, ou génétique inverse, est l'outil majeur de la génétique humaine, mais l'art est difficile… Le choix initial de la maladie à étudier est un gambit, dont il faut savoir limiter la prise de risque. C'est une quête longue, fastidieuse, reposant sur la participation volontaire d'un ensemble de familles et le travail acharné d'un groupe de chercheurs. Les premières maladies neurologiques (Huntington, neurofibromatose et sclérose tubéreuse de Bourneville, dégénérescences spino-cérébelleuses, myotonie de Steinert) ont été clonées ces cinq dernières années. La découverte du gène responsable d'une maladie permet, en principe, de mieux comprendre les relations entre l'homme et son ADN…

Cloner, pourquoi ?

Le clonage positionnel, ou découverte du gène responsable d'une maladie humaine par génétique inverse, a été utilisé avec succès pour la première fois en 1986, avec la découverte du gène de l'agranulocytose septique, déficit immunitaire sévère du garçon, localisé en Xp21.1 *(Royer Pakora, 1986)*. Cette avancée fondamentale, gratifiée d'un article dans la revue prestigieuse *Nature*, est le fruit de la recherche sur plus de cinq ans d'un laboratoire spécialisé en immunologie. Actuellement, la puissance des outils de biologie moléculaire et le perfectionnement de la carte du génome humain permettent de réduire ce temps, mais le clonage positionnel reste une prouesse technique saluée par les pairs.

Clonage positionnel, prérequis

L'initiation d'un clonage positionnel est un choix stratégique, toujours mûrement réfléchi, qui entraîne un laboratoire complet dans une quête du Graal longue, coûteuse et risquée. La décision prend en compte les critères suivants :

– la précision du phénotype clinique : ou comment être sûr que toutes les familles étudiées sont bien atteintes de la même maladie. Une hétérogénéité clinique rend la recherche ardue et sujette à résultats discordants ;

– le nombre potentiel des familles : une maladie extrêmement rare a plus de difficulté à bénéficier de ce type d'approche ;

– la collection de l'ensemble des membres d'une famille : pour une famille donnée, la précision de la liaison (ou linkage) repose sur la coopération de **tous** les membres de la famille, atteints et non atteints (indispensable à l'étude), ce qui conduit à une collecte de prélèvements étendue, souvent manuelle (avec le généticien et l'infirmière dans la voiture, sur les routes de France). Il est donc crucial que l'ensemble des membres des familles aient envie de contribuer à la recherche, dans un but curatif ou pour améliorer le diagnostic prénatal des maladies très sévères ;

– la possibilité de distinguer les sujets atteints et non atteints : dans les maladies génétiques dominantes à pénétrance variable (expression plus ou moins visible de la maladie quand on est porteur du gène muté), le risque majeur est de considérer comme non atteinte une personne porteuse du gène sans expression clinique. L'examen clinique par un clinicien averti, qui reconnaîtra tous les petits signes de la maladie, est donc important ;

– le nombre des chercheurs et le budget du laboratoire : ce type d'étude nécessite au moins quatre ou cinq chercheurs et techniciens sur le sujet, pour réussir… en moins de cinquante ans, et une connaissance approfondie du domaine pour savoir vers quoi on se lance ! *(tableau).*

Techniques utilisées

Génome scan

La première étape est de rechercher quelle partie du génome est toujours transmise chez tous les patients atteints, et transmise aléatoirement pour les sujets sains. L'étude de cette liaison maladie/localisation subchromosomique est effectuée par l'analyse systématique de la transmission des marqueurs microsatellites (séquences répétées réparties sur l'ensemble du génome) au sein de chaque famille, pour trouver une liaison génétique entre le locus supposé de la maladie et un ou plusieurs microsatellites. Le principe est simple mais très répétitif : regarder dans chaque famille quel microsatellite est toujours transmis d'un parent atteint à un enfant atteint.

Maladie dominante

Par exemple, dans la famille A, le microsatellite D1S450 (chromosome 1) est présent sous les formes 32 et 17 répétitions chez le

* *Service de pédiatrie néonatale, hôpital intercommunal, Créteil.*

Maladies	Localisation du gène	Protéine	Fonction	Type d'altération du gène
Malformations cérébrales				
Holoprosencéphalie	7q36	Sonic Hedgehog	Inducteur neuronal	Délétions, mutations ponctuelles
Holoprosencéphalie	13q32	ZIC2	Inducteur neuronal	Délétions, mutations ponctuelles
Holoprosencéphalie	2p21	SIX3	Inducteur neuronal	Délétions, mutations ponctuelles
Hydrocéphalie liée à l'X	Xq28	L1CAM	Protéine de l'adhésion cellulaire	Délétions, mutations ponctuelles
Anomalie de migration neuronale				
Agyrie-pachygyrie, Miller Dieker	17p13,3	LIS1	Rôle dans la liaison aux microtubules?	Délétions, mutations ponctuelles
Double cortex, lissencéphalie liée à l'X	Xq22,3	Double Cortine	Stabilisateur de microtubules	Mutations ponctuelles
Hétérotopies laminaires périventriculaires	Xq28	Filamine 1	Protéine se liant à l'actine	Duplications, mutations ponctuelles
Maladies métaboliques				
CDG de type IA	16p13,2	PMM2 Phosphomannomutase	Glycosylation des protéines	Mutations ponctuelles
Épilepsies				
Épilepsie bénigne du nouveau-né	20q13,3	KCNQ2	Canal potassique	Mutations ponctuelles
Épilepsie familiale idiopathique	8q24	KCNQ3	Canal potassique	Mutations ponctuelles
Maladie de Lafora	6q24	EPM2A	Tyrosine phosphatase	Délétions, mutations ponctuelles
Phacomatoses				
Neurofibromatose de type 1	17q11,2	Neurofibromine	Gène suppresseur de tumeur	Délétions, duplications, mutations ponctuelles
Neurofibromatose de type 2	22q11	Merline	Gène suppresseur de tumeur	Délétions, mutations ponctuelles
Sclérose tubéreuse de Bourneville	9q34	Hamartine	Gène suppresseur de tumeur	Délétions, mutations ponctuelles
Sclérose tubéreuse de Bourneville	16p13,3	Tubérine	Gène suppresseur de tumeur	Délétions, mutations ponctuelles
Maladie de von Hippel-Lindau	3p25	VHL	Gène suppresseur de tumeur	Délétions, mutations ponctuelles
Atteinte des noyaux gris centraux				
Dystonie généralisée familiale	9q34	DYT1 Torsine A	Protéine chaperonne ?	Mutations ponctuelles
Maladie de Wilson	13q14,3	ATP7B	Transporteur du cuivre	Mutations ponctuelles
Démences				
Alzheimer 1, familial à début précoce	21q21	APP	Précurseur de l'amyloïde	Mutations ponctuelles
Alzheimer 2, tardif	19q13.2	Apolipoprotéine E4	Lipoprotéine	Homozygotie apo E (E4/E4)
Alzheimer 3, familial à début précoce	14q24.3	Préséniline 1	Protéine transmembranaire	Mutations ponctuelles
Alzheimer 4, familial à début précoce	1q31.42	Préséniline 2	Protéine transmembranaire	Mutations ponctuelles
Maladie de Huntington	4p16,3	Huntingtine	Inconnu	Expansion de répétions de tri-nucléotides
CADASIL	19p13,1	Notch3	Protéine du signal intercellulaire	Mutations ponctuelles
Maladie de Kennedy	Xq13	Récepteur aux androgènes	Rôle inconnu sur la corne antérieure	Expansion de répétions de tri-nucléotides
Creutzfeldt-Jakob familial, insomnie familiale fatale	20p12	PRNP, protéine du prion	Protéine chaperonne	Mutations ponctuelles
Neuropathies périphériques				
Maladie de Charcot-Marie-Tooth de type 1A	17p11,2	PMP22	Protéine de la myéline	Duplications, mutations ponctuelles
Maladie de Charcot-Marie-Tooth de type 1B	1q22-23	P0	Protéine de la myéline	Mutations ponctuelles
Maladie de Charcot-Marie-Tooth liée à l'X	Xq13	Connexine 32	Contact intercellulaire	Délétions, mutations ponctuelles
Dystrophies musculaires				
Myotonie de Steinert	19q13,2	Myotonine	Protéine-kinase	Expansion de répétions de tri-nucléotides
Emery Dreifuss	9q34	Émérine	Protéine membranaire, rôle?	Délétions, mutations ponctuelles
Maladies Neurodégénératives				
Sclérose latérale amyotrophique	21q22,1	Superoxyde dismutase	Épuration de radicaux libres	Mutations ponctuelles
Ataxie de Friedriech	9q13	X25 frataxine	Échangeur fer-soufre mitochondrial	Expansion de répétions de tri-nucléotides
Paraplégie spastique familiale dominante	2p21	Spastine	Protéine chaperonne	Mutations ponctuelles
Retard mental				
Syndrome de l'X fragile	Xq27	FMR1	Protéine liant l'ARN	Expansion de répétions de tri-nucléotides
Anomalies vasculaires				
Angiomes caverneux héréditaires	7q	CCM1	Protéine de signal cellulaire	Mutations ponctuelles

Tableau I. Tableau non exhaustif des principales maladies neurologiques dont le gène a été localisé et cloné par génétique inverse ; en perpétuel remaniement, tant il est vrai qu'il y a maintenant plusieurs gènes par maladie et plusieurs phénotypes par gène cloné !

père atteint, qui a transmis à ses douze fils atteints le microsatellite D1S450 sous sa forme 32 répétitions, ce qui n'est pas le fruit du hasard ! On peut espérer voir confirmer cette transmission préférentielle dans une autre famille B, pour ce même microsatellite D1S450, avec la transmission d'un allèle 18 répétitions de la mère atteinte à ses dix enfants atteints, et dans la famille C… L'observation est particulièrement facilitée dans les cas de grandes familles comprenant de nombreux malades *(figure 1)*.

Maladie récessive

Le principe est le même, en comparant la transmission d'un microsatellite d'un parent porteur obligatoire (hétérozytgote), à un enfant atteint (homozygote). La consanguinité favorise ce type d'étude : c'est l'*Homozygoty Mapping*.

Cette étape est longue et décourageante, puisqu'il faut tester au pire 400 à 500 microsatellites pour localiser la maladie sur une région chromosomique. La méthode du Lod-score permet de tester statistiquement cette association maladie/microsatellite, et un résultat est probant si le Lod-score est supérieur à 3 (probabilité d'une liaison significative maladie/microsatellite supérieure à 1 sur 1 000).

Pour raccourcir cette étape, tout mouton à cinq pattes est recueilli précieusement :

– association de la maladie à une translocation chromosomique équilibrée (étude des points de cassure) ;

– association morbide de plusieurs maladies différentes, révélatrice d'une délétion chromosomique.

Marche sur le chromosome

Quand le Lod-score permet de suspecter une région chromosomique donnée, la cartographie physique de la région peut alors être entreprise. Et cette étape est celle du saut quantique, passant de l'infiniment visible (chromosomique) à l'infiniment invisible (le triplet de bases nucléiques). Et, à côté de cela, trouver une aiguille dans une meule de foin, c'est bien connu pour être beaucoup plus facile !

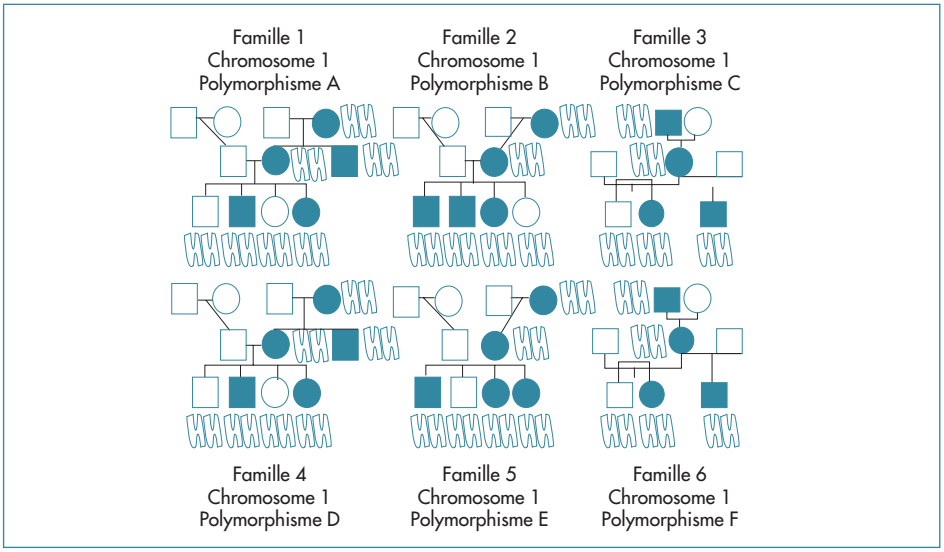

Figure 1. Localisation d'un locus morbide : mais quelle est donc la région chromosomique toujours transmise d'un parent atteint à un enfant atteint ? Familles nombreuses souhaitées…

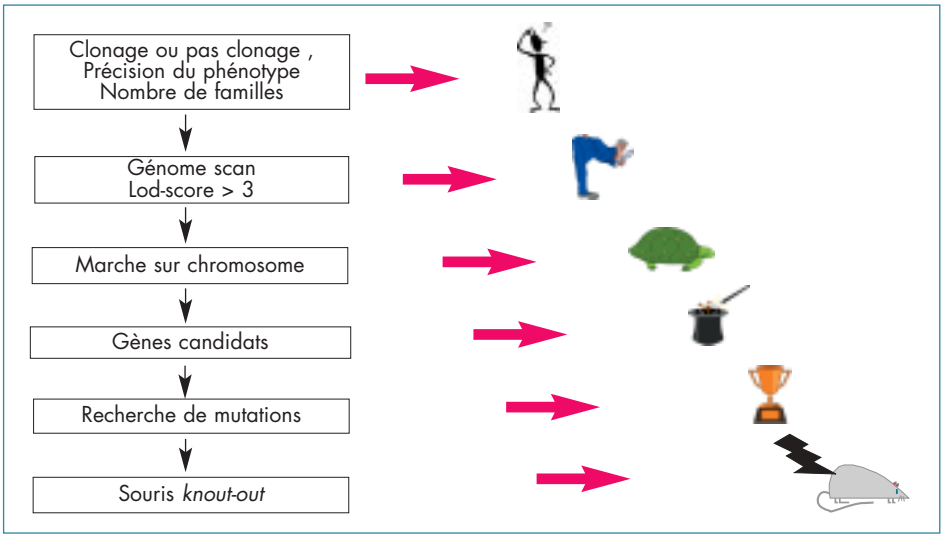

Figure 2. Clonage positionnel, mode d'emploi…

En fonction de la localisation retrouvée, des informations génétiques précises ont déjà pu être recueillies dans la région (ainsi, le chromosome X qui est multiexploré), accélérant cette étape.

La technique de "marche sur le chromosome" consiste à s'appuyer sur des YAC, BAC ou PAC (chromosomes artificiels de levure, bactérie ou phage qui contiennent un fragment de génome humain) chevauchant la région d'inté-

rêt. En effet, un ensemble de ces chromosomes artificiels parcourant l'intégralité du génome a été fabriqué grâce à l'action de plusieurs équipes de recherche *(Cohen, 1993)* et est en perpétuel enrichissement. Une fois le ou les YACS sélectionnés, il faut aller à la pêche au gène, soit en séquençant bestialement l'ensemble du fragment de génome, soit en recherchant des séquences qui évoquent une séquence codante (éléments spécifiques d'un gène exprimé), ou *exon trapping.*

Gènes candidats

Dans certains cas, des gènes peuvent avoir préalablement été localisés dans cette région d'intérêt. La séquence de ces gènes sera systématiquement testée, surtout si la fonction potentielle du gène peut avoir une liaison avec le phénotype (ainsi l'ostéogenèse imparfaite et les gènes du collagène…). Cette méthode permet alors de court-circuiter la phase d'exploration systématique de la région.

Recherche de mutations

Une fois la séquence codante du gène découverte, c'est la course aux premières mutations invalidant le gène, responsables de sa perte de fonction complète ou partielle, qu'elles soient des insertions, des délétions ou des mutations non-sens. Il faut alors publier très vite, par crainte d'un scoop publié par l'équipe adverse (car il y a toujours une équipe adverse, ce qui stimule la paranoïa des équipes de recherche).

Validation des mutations

La validation des mutations de type faux-sens (modifiant simplement la séquence du gène sans le détruire complètement) doit être démontrée par la constatation d'une anomalie de la fonction de la protéine codée par ce gène permettant ainsi de les différencier d'un simple polymorphisme de séquence. Cette étape est également longue et fastidieuse, car elle implique de transfecter (introduire la séquence du gène porteuse de la mutation) dans des cellules en culture pour vérifier que cette anomalie est bien responsable d'une non-fonctionnalité de la protéine produite.

Souris *knock-out*

La dernière étape de cette course au gène consiste à reproduire la maladie chez la souris, en remplaçant le gène normal de la souris par un gène non fonctionnel (porteur d'une mutation perturbant la fonction normale de la protéine). Le phénotype présenté par la souris est au mieux proche de la maladie découverte chez l'homme (comme les souris sans yeux obtenues après inactivation du gène fondamental de la formation de l'œil, le gène PAX6). Mais il est vrai que l'homme n'est pas la souris, et des différences fondamentales ont déjà été démontrées (souris porteuses d'une mutation du gène de la dystrophine sans myopathie, mutations délétères chez l'homme à l'état hétérozygote et sans expression chez la souris).

Innovations technologiques

La dernière technique actuellement développée est celle des *DNA MicroChips*, permettant l'étude de multiples séquences géniques en même temps, ouvrant la porte à des études complexes, et pouvant faire rêver d'une analyse "dégénérescence spinocérébelleuse" (plus de 12 gènes étudiés), "retard mental avec dysmorphie" (au moins 50 gènes actuellement clonés), ou "retard mental sans dysmorphie" (plusieurs dizaines de gènes à terme). Cette technique permettrait même l'analyse de plusieurs centaines de polymorphismes et mutations.

Recette du clonage positionnel *(figure 2).*

Conclusion

La technique de clonage positionnel est actuellement extrêmement productive, permettant le clonage de plusieurs gènes fondamentaux tous les mois. Récemment, des gènes de l'épilepsie infantile (KCNQ2 et KCNQ3, gènes de canaux potassiques responsables d'épilepsie bénigne néonatale, et encore EPM2A, gène de l'épilepsie myoclonique progressive, ou maladie de Lafora) ont été clonés, ainsi que des gènes du fonctionnement vasculaire cérébral (KRIT1, gène des angiomes caverneux familiaux, NOTCH3, gène du CADASIL, démence vasculaire avec infarctus cérébraux). Malheureusement, la découverte de ces multiples gènes ne permet pas toujours de comprendre le rapport entre la fonction du gène (plus ou moins bien étudiée) et l'expression clinique chez l'homme. Cette phase de compréhension de la corrélation génotype/phénotype est ardue et nécessite des investigations complexes. Seront-elles l'ultime étape de la recherche ?

Mots clés. Clonage positionnel – Génétique inverse.

Pour en savoir plus

- *Charlier C, Singh NA, Ryan SG et al. A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family. Nature Genetics 1998 ; 18(1) : 53-5.*
- *Cohen D, Chumakov I, Weissenbach J. A first-generation physical map of the human genome. Nature 1993 ; 366 : 698-701.*
- *Joutel A, Corpechot C, Ducros A et al. NOTCH3 mutations in CADASIL, an hereditary adult-onset condition causing stroke and dementia. Nature 1996 ; 383 : 707-10.*
- *Kaplan JC, Delpech M. Biologie moléculaire et Médecine. Médecine-Science, Presses Flammarion 1992.*
- *Kitzis A, Warren P, Kaplan JC. Génétique inverse et mucoviscidose. Médecine/Sciences Presses Flammarion 1988 ; 4 :151-6*
- *Laberge-Le Couteulx S, Jung HH, Labauge P et al. Truncating mutations in CCM1, encoding KRIT1, cause hereditary cavernous angiomas. Nature Genetics 1999 ; 23(2) : 189-93.*
- *Minassian BA, Lee JR, Herbrick JA et al. Mutations in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy. Nature Genetics 1998 : 20(2) : 171-4.*
- *Royer-Pokora B, Kunkel LM, Monaco AP et al. Cloning the gene for an inherited human disorder–chronic granulomatous disease, on the basis of its chromosomal location. Nature 1986 ; 322 : 32-8.*
- *Singh NA, Charlier C, Stauffer D et al. A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns. Nature Genetics 1998 ; 18(1) : 25-9.*